Soil and Tillage Research, Volume 132, August 2013, Pages 21–29

ON-LINE MEASUREMENT OF SOIL PROPERTIES WITHOUT DIRECT SPECTRAL RESPONSE IN NEAR INFRARED SPECTRAL RANGE

Omar Marín González[1], Boyan Kuang[2], Mohammed Z. Quraishi[2], Miguel Ángel Muñoz García[1], Abdul Mounem Mouazen[2]

[1] Rural Engineering Department, Electrotechnical Section, EUITAgrícola, UPM Avenida Complutense, s/n; 28040 Madrid, Spain

[2] Environmental Science and Technology Department, Cranfield University, Bedfordshire, MK43 0AL, United Kingdom

Corresponding author: o.marin@upm.es; Phone: +34 689616130; Fax: +34 915449983

Abstract

So far, the majority of reports on on-line measurement considered soil properties with direct spectral responses in near infrared spectroscopy (NIRS). This work reports on the results of on-line measurement of soil properties with indirect spectral responses, e.g. pH, cation exchange capacity (CEC), exchangeable calcium (Ca_{ex}) and exchangeable magnesium (Mg_{ex}) in one field in Bedfordshire in the UK. The on-line sensor consisted of a subsoiler coupled with an AgroSpec mobile, fibre type, visible and near infrared (vis-NIR) spectrophotometer (tec5 Technology for Spectroscopy, Germany), with a measurement range of 305 - 2200 nm to acquire soil spectra in diffuse reflectance mode. General calibration models for the studied soil properties were developed with a partial least squares regression (PLSR) with one-leave-out cross validation, using spectra measured under non-mobile laboratory conditions of 160 soil samples collected from different fields in four farms in Europe, namely, Czech Republic, Denmark, Netherland and UK. A group of 25 samples independent from the calibration set were used as independent validation set. Higher accuracy was obtained for laboratory scanning as compared to on-line scanning of the 25 independent samples. The prediction accuracy for the laboratory and on-line measurements was

28 classified as excellent/very good for pH (RPD = 2.69 and 2.14 and r^2 = 0.86 and 0.78,
29 respectively), and moderately good for CEC (RPD = 1.77 and 1.61 and r^2 = 0.68 and
30 0.62, respectively) and Mg_{ex} (RPD = 1.72 and 1.49 and r^2 = 0.66 and 0.67,
31 respectively). For Ca_{ex}, very good accuracy was calculated for laboratory method (RPD
32 = 2.19 and r^2 = 0.86), as compared to the poor accuracy reported for the on-line
33 method (RPD = 1.30 and r^2 = 0.61). The ability of collecting large number of data points
34 per field area (about 12800 point per 21 ha) and the simultaneous analysis of several
35 soil properties without direct spectral response in the NIR range at relatively high
36 operational speed and appreciable accuracy, encourage the recommendation of the
37 on-line measurement system for site specific fertilisation.
38
39 **1 Introduction**

40 During the last decade visible near infrared spectroscopy (vis-NIRS) is being
41 increasingly used to detect soil properties, with variable accuracy depending on several
42 factors (Kuang et al., 2012). The vis-NIRS is a simple, non-destructive and rapid
43 technique, needs no sample preparation for field applications, and can be used for the
44 laboratory, *in situ* (Viscarra Rossel et al., 2006a), and on-line measurements (Mouazen
45 et al., 2005). Furthermore, in some cases accuracy obtained is high and very similar to
46 that of conventional procedures (Viscarra Rossel et al., 2001). The technique allows
47 the assessment of primary properties with direct spectral responses, which are directly
48 affected by combinations and overtones of fundamental vibrations for organic
49 functional groups and water, particle size, and surface properties (Chang et al., 2001).
50 In a review paper, Stenberg et al. (2010) defined organic carbon (OC), total carbon
51 (TC), moisture content (MC) and clay minerals to have direct spectral response in the
52 NIR spectroscopy. In this context, it is possible to detect clay type and content, MC, OC
53 and TC with high accuracy (Volkan et al., 2010). Other soil properties without direct
54 spectral absorption features in the vis-NIR range (secondary properties) can be also

55 measured with good to moderate accuracy due to co-variation with one or more
56 primary properties (Stenberg et al., 2010). For both soil categories, colour plays an
57 important role in enhancing measurement accuracy. Good results have been reported
58 for cation exchange capacity (CEC), pH, extractable calcium (Ca_{ex}) and extractable
59 magnesium (Mg_{ex}) under non-mobile laboratory and *in situ* conditions but
60 underperformed those for properties with direct spectral responses in the NIR range,
61 e.g. OC, TC, MC and clay content (Kuang et al., 2012). According to the literature,
62 sodium (Na_{ex}) and potassium (K_{ex}) are among the most difficult properties to be
63 measured with the NIR spectroscopy (Malley et al., 1999, Chang et al., 2001, Zornoza
64 et al., 2008, Pirie et al., 2005, Dunn et al., 2002, Shepherd & Walsh, 2002, Islam et al.,
65 2003, Volkan et al., 2010; Mouazen et al., 2006). For the same soil property, laboratory
66 vis-NIR methods achieved higher accuracy as compared to measurement under field
67 soil conditions, particularly with on-line vis-NIR sensors (Stenberg et al., 2010; Kuang
68 et al., 2012).
69 The prediction accuracy achieved so far with the on-line vis-NIRS sensors available
70 today (Shibusawa et al, 2001; Mouazen, 2006; Christy, 2008) might be sufficient for
71 many applications in precision agriculture, since spatial and temporal variation of soil
72 properties is large relative to the precision of measurement (Shepherd and Walsh,
73 2002). However, similar to laboratory and *in situ* measurements, Kuang et al, (2012)
74 concluded that the best accuracy is achieved for soil properties with direct spectral
75 response, which is probably the reason why researchers using on-line vis-NIRS
76 sensors have focused mainly on soil properties with direct spectral responses (Shonk
77 et al., 1991; Mouazen et al., 2005; Bricklemyer et al., 2010; Munoz and Kravchenko,
78 2011, Knadel et al., 2011). Mouazen et al. (2007) showed potential success for the on-
79 line vis-NIRS measurement of extractable and available P and pH, without proving the
80 accuracy to be of quantitative meaning. Although Mouazen et al. (2007) demonstrated
81 spatial similarities between measured and on-line predicted pH, no robust conclusions
82 on accuracy of measurement could be drawn. Later Mouazen et al. (2009) conducted

on-line measurement of available P with remarkable accuracy (RPD = 1.42; r^2 = 0.62). To our knowledge none of the previous studies has reported on the on-line measurement of CEC, Ca_{ex} and Mg_{ex}. In addition, no comparison was made between laboratory and to on-line measurement of the named soil properties.

The aim of this paper is to evaluate the performance and accuracy of on-line measurement of soil properties without direct spectral responses in the NIR spectroscopy range, namely, CEC, pH, Ca_{ex} and Mg_{ex}. It also aims to compare the prediction accuracy of these properties based on soil spectra collected under on-line measurement conditions with those collected under laboratory non-mobile conditions. Calibration models developed for several farms in Europe will be used to validate the on-line measurement of these properties in one selected field in the UK.

2 Materials and methods

2.1 Soil samples

A total of 140 soil samples were used to develop general calibration models for the prediction of pH, CEC, Ca_{ex}, and Mg_{ex}. These soil samples were collected from different fields in six farms in four different European countries (Fig. 1 and Table 1). A total of 25 samples were collected from one field in Mespol Medlov, A.S. farm (Czech Republic), 20 samples from two fields in Bramstrup Estate farm (Denmark), 23 samples from one field in Wageningen University (the Netherland), 25 samples from one field in Ely farm (Cambridgeshire, UK), 17 samples from five different fields in Silsoe experimental farm (Bedfordshire, UK) and 30 samples from fields 2 and 3 in Duck End farm (Bedfordshire, the UK) (Table 2). Bulked samples were collected from the upper soil layer (0-30 cm) in the spring of 2008 in the Mespol Medlov, A.S. farm and in the spring of 2009 in Bramstrup Estate farm (Kuang and Mouazen, 2011). For the remaining farms, soil samples were collected from the bottom of 15 cm deep trenches opened by a subsoiler during the on-line vis-NIR measurement in the autumn of 2010

in Wageningen University, summer of 2009 in Silsoe experimental farm, spring of 2011 in Ely farm, and summer of 2011 in Duck End farm. These fields correspond to a large diversity of soil textures, crops and landscapes (Table 1).

In order to validate the calibration models developed, on-line measurement was carried out in field 1 in Duck End Farm (Table 2). During this measurement a total of 45 samples were collected for model development and validation. About 44 % of these 20 samples were added to the general calibration set (140 + 20 = 160 samples) to develop the general calibration models of the selected four properties, whereas the remaining 25 samples were used as validation set for the validation of the general calibration models based on non-mobile (laboratory) and on-line measured spectra of these 25 samples.

Approximately 200 g of each soil sample was kept deep frozen (-18 °C) until testing. Before analysis, each sample was defrosted, carefully mixed and divided into two portions, with one used for chemical analysis, and the other used for optical measurement. Samples were stored in plastic bags at 4 °C during the analysis to avoid losing humidity.

2.2 Laboratory measurement of chemical soil properties

Laboratory analyses of pH, CEC and base saturation, were performed at the soil laboratory of the National Soil Resource Institute, School of Applied Sciences (NR-SAS) of Cranfield University (Bedfordshire, the UK) using their standard procedures. Soil pH was measured in a 1:5 soil:H_2O suspension following the BS ISO 10390 (2005) for the determination of soil pH. In order to determine the CEC and exchangeable cations (Na_{ex}, K_{ex}, Ca_{ex}, Mg_{ex}), the air-dried soil samples were first saturated with respect to barium by adding 30 ml of barium chloride solution (Reagent Production Unit (RPU) 10), after which 30 ml excess of 0.02 mol/l magnesium sulphate (RPU 11) was added. This makes all barium present in the solution as well as adsorbed in

137 exchangeable sites to precipitate in the form of highly insoluble barium sulphate. As a
138 result, the exchangeable sites are occupied by magnesium. The surplus of magnesium
139 was determined by atomic absorption following the NR-SAS SOP 42/Version 1, based
140 on the BS 7755 section 3.12 (1996), for the determination of the potential cation
141 exchange capacity and exchangeable cations using barium chloride buffered at
142 pH=8.1, which is identical to ISO 13536 (1995) (Table 3). Base saturation was
143 calculated as the equivalent sum of major base cations (Ca_{ex}, Mg_{ex}, K_{ex} and Na_{ex})
144 percentage of CEC (Chodak et al, 2004).

145

146 **2.3 Optical measurement**

147 **2.3.1 Optical measurement in laboratory**

148 For each soil sample a certain amount of soil was mixed up in a glass bowl. Stones
149 and plant residues were removed at this point. Three small cups of 1 cm deep and 3.6
150 cm in diameter were filled up with the same soil sample. The surface of the samples
151 was softly pressed and smoothed down with a spatula, simulating the effect of the
152 subsoiler smoothing of soil beneath the chisel during on-line measurement, which
153 increases the signal to noise ratio (Mouazen et al., 2005).
154 An AgroSpec mobile, fibre type, vis-NIR spectrophotometer (Tec5 Technology for
155 Spectroscopy, Germany) with a measurement range of 305 - 2200 nm was used to
156 measure soil spectra in diffuse reflectance mode. A 100 % ceramic was used as the
157 white reference, which was scanned once every 30 minutes. Optical scanning was
158 conducted on non-treated, fresh soil samples to simulate field measurement conditions.
159 A total of ten scans were performed for each of the three plates prepared for each soil
160 sample. The resulted thirty spectra were averaged into one spectrum for each sample
161 (Kuang and Mouazen, 2011). This averaged spectrum was used for spectra pre-

162 treatment and model development. All calibration (160) and validation (25) samples
163 were scanned in the laboratory.
164

165 **2.3.2 On-line vis-NIR measurement**

166 The on-line spectra were collected along parallel measurement lines in the field 1 in the
167 Duck End Farm in the UK (Fig. 2). Detailed information about this field is shown in
168 Table 4. The same vis-NIR spectrophotometer used for laboratory measurement
169 (AgroSpec) was used for on-line measurement. The on-line sensor developed by
170 Mouazen (2006) was used to carry out on-line measurement. The subsoiler makes a
171 trench in the soil, whose bottom is smoothened by the subsoiler itself, due to the
172 downwards vertical forces. The optical probe protected in a steel lens holder was
173 appended to the backside of the subsoiler chisel to measure soil spectra in diffuse
174 reflectance mode from the smoothened bottom of the 15 cm deep trench created by
175 the subsoiler chisel. The subsoiler retrofitted with the optical probe was attached to a
176 frame, which was mounted onto the three point linkage of a tractor (Fig. 3). The
177 spectrophotometer was IP 64 protected for harsh working environments. A deferential
178 global positioning system (DGPS) (EZ-Guide 250, Trimble, USA) was used to record
179 the position of on-line measured spectra with sub-meter accuracy. A Panasonic semi-
180 rugged laptop was used for data logging and communication. The spectrophotometer,
181 laptop and DGPS were powered by the tractor battery.
182

183 **2.4 Pre-processing of spectra**

184 Several spectra pre-processing were tested and the best performing one was kept. The
185 selection criteria of any pre-processing were the largest coefficient of multiple
186 determination (r^2) and residual prediction deviation (RPD), which is the ratio of standard
187 deviation (S.D.) of the prediction data set to root mean square error of prediction

188 (RMSEP) and the smallest RMSEP. Spectra pre-processing and establishment of
189 calibration models of different properties were done by Unscrambler 7.8 software
190 (Camo Inc., Oslo, Norway). To remove noise at edges of spectra, soil spectrum was
191 first abridged to 400 – 2100 nm for each sample. After noise was removed, spectra
192 were reduced by averaging three successive wavelengths in the visible range (400 -
193 1000 nm) and six successive wavelengths for the NIR region (1000 - 2100 nm).
194 Averaging over wavelengths was used to decrease the number of wavelengths and to
195 smooth the spectrum (Nicola et al., 2007). Maximum normalization was followed, which
196 is typically used to get all data to approximately the same scale, or to get a more even
197 distribution of the variances and the average values. This method attempted to remove
198 the effects of scattering by linearising each spectrum to some 'ideal' spectrum of the
199 sample, which, in practice, corresponds to the average spectrum (Nicola et al., 2007).
200 The maximum normalization 'polarizes' the spectra. The peaks of all spectra with
201 positive values are scaled to +1, while spectra with negative values are scaled to -1.
202 The peaks of these spectra were scaled to +1, since all soil spectra in this study had
203 positive values (Mouazen et al., 2005). The maximum normalization was selected
204 because it provided better results for all properties considered compared with other
205 pre-processing tested. Spectra were subsequently subjected to Savitzky–Golay first
206 derivation (Martens and Naes, 1989). This transformation procedure generally
207 intensifies the absorption characteristics indicative of soils properties, and diminishes
208 variation among spectra (Volkan et al., 2010). This method enabled the computation of
209 the first or higher-order derivatives, including a smoothing factor, which determines
210 how many adjacent variables should be used to estimate the polynomial approximation
211 used for derivatives. A second-order polynomial approximation was selected with a 2:2
212 smoothing factor. A 2:2 Savitzky-Golay smoothing was carried out after the first
213 derivative to remove random noise from spectra (Kuang and Mouazen, 2011).
214

2.5 Establishment of calibration models

The pre-processed spectra and the results of laboratory chemical analyses were used to develop calibration models for all studied properties. Before partial least squares regression (PLSR) analysis, the 160 calibration samples (Tables 1 & 2) were divided into calibration (85 % of samples) and prediction sets (15 % of samples). Further validation of developed models was done using the 25 samples collected in the field 1 in Duck End farm, which was designated as independent validation set.

To develop calibration models, especial attention should be paid in the selection of calibration and validation sets. The distribution and size of the calibration data set must cover the entire range of concentration of a soil property examined. To build a robust calibration model, validation of model developed has to be carried out using validation samples, which were not used for model development. This means that validation samples of the independent validation set (25 samples collected from the field 1 in the Duck End farm) should not be used for the development of calibration models in cross-validation, avoiding this way any influence in the prediction capacity of the model selected and the consequent overestimation (Brown et al., 2005). The same samples of calibration and validation were selected for each property. A previous threshing of samples was executed to get the wider range of values and spectra variation in the calibration set. Calibration and validation samples were then selected randomly. Performing this way a calibration data set was set to cover the whole range of concentration for the different soil properties examined.

The most used multivariate methods to develop calibration models are based on linear regressions. Mainly, stepwise multiple linear regression (SMLR), principal component regression (PCR), and PLSR (Stenberg et al., 2010) are used. The PLSR with leave-one-out cross-validation was carried out using Unscrambler 7.8 software (Camo Inc., Oslo, Norway) to generate the calibration models relating soil independent variables (wavelengths) of the diffuse reflectance spectra to each soil parameter. PLSR performs

particularly well, compared with other multivariate statistical methods, when there is a high dimensional correlation between variables, which is the case for soil spectral data (Volkan et al, 2010). Also, PLSR is favoured because it requires fewer components to explain the variance in the response, due to the relation that this method establishes between response and predictor variables, and its results are more interpretable (Stenberg et al., 2010)

The number of latent variables for a model was determined by examining a plot of leave-one-out cross-validation residual variance against the number of latent variables obtained from PLSR. The latent variable of the first minimum value of residual variance was selected (Brown et al., 2005). The residual sample variance and predicted vs. measured plots were assessed for outliers determination after running the PLSR. Outliers may be induced by typing errors, file transfer, interface errors, sensor malfunctions and fouling, poor sensor calibration, bad sampling or sample presentation, etc. (Nicola et al., 2007). Samples located individually far from the zero line of residual variance together with a far position from the trend line in the predicted vs. measured plot were considered as outliers and excluded from the analysis. A maximum of 5 % of the entire samples was accepted as the maximum number of outliers to be removed (Kuang and Mouazen, 2011).

2.6 Performance assessment of calibration models

Prediction accuracy of a PLSR model was determined by the RMSEP, r^2 and RPD. The criteria adopted for RPD classification (Viscarra Rossel et al., 2006b) was as follows: RPD < 1.0 indicates very poor model/predictions and their use is not recommended; RPD between 1.0 and 1.4 indicates poor model/predictions where only high and low values are distinguishable; RPD between 1.4 and 1.8 indicates fair or moderately good model/predictions which may be used for assessment and correlation; RPD values between 1.8 and 2.0 indicates good model/predictions where quantitative predictions

are possible; RPD between 2.0 and 2.5 indicates very good, quantitative model/predictions, and RPD > 2.5 indicates excellent model/predictions. This classification system was adopted in this study.

3 Results and discussion

For robust modelling of the vis-NIR spectral data, the selection of calibration set should be carefully done so that to be representative of the samples used for validation. The range for sample concentrations in the prediction (Table 5) and independent validation (Table 6) sets is smaller and within the corresponding range for the calibration set (Table 5) for all properties. This confirms that variation in the prediction and independent validation sets are accounted for in the calibration set.

Since samples of Ely field were of excessively high exchangeable Ca_{ex} and Mg_{ex} values, the 25 samples from Ely field were disposed during the development of the calibration models for these two properties, as these might negatively affect the prediction accuracy of the corresponding models.

3.1 Evaluation of general calibration models

Examining the results of the prediction set (Table 7), reveals that best results are achieved for Mg_{ex} (r^2 = 0.88; RPD = 2.55) and pH (r^2 = 0.86; RPD = 2.37). Less accurate estimations were obtained for CEC (r^2 = 0.72; RPD = 1.70) and Ca_{ex} (r^2 = 0.76; RPD = 1.87). According to the classification based on RPD proposed by Viscarra Rossel et al. (2006b), the prediction accuracy in the prediction set is excellent and very good for Mg_{ex} and pH, respectively, good for Ca_{ex} and moderately good for CEC. These results are in line with those found in the literature. Comparing with other studies reported in the literature under laboratory measurement condition (Shepherd and Walsh, 2002; Cohen et al., 2005; Mouazen et al., 2006; Viscarra Rossel and Behrens.,

295 2010), the model performance for pH in the prediction set is among the best models (r^2
296 = 0.50 - 0.97; RMSEP = 0.04 - 1.43; RPD = 0.57 - 2.39). Similar conclusion can be
297 drawn for the prediction of Mg_{ex}, with overall results reported in the literature for r^2 of
298 0.53 - 0.91; RMSEP of 0.03-38.36 cmol kg-1 and RPD of 0.48 - 2.54 (Cozzolino and
299 Moron, 2003; Groenigen et al., 2003; Udelhoven et al., 2003; Wetterlind et al., 2010).
300 Almost the same results as those achieved in this study were reported by Dunn et al.
301 (2002), when analysing pH for the top soil of 0 - 10 cm (r^2 = 0.80, RPD = 2.3) and Mg_{ex}
302 (r^2 = 0.85, RPD = 2.7). The models for the other two properties are less accurate, as
303 compared with other values reported in the literature for Ca_{ex} (r^2 = 0.07-0.95; RMSEP =
304 0.66 - 52.90 cmol kg^{-1}; RPD = 0.60-2.75) (e.g. Cozzolino and Moron, 2003; Cohen et
305 al., 2005; Mouazen et al., 2006; Zornoza et al., 2008) and CEC (r^2 = 0.13 - 0.90;
306 RMSEP = 1.22 - 10.43 cmol kg^{-1}; RPD = 0.60 - 2.7) (e.g. Ben-Dor and Banin, 1995;
307 Chang et al., 2001; Mouazen et al., 2006; Brown et al., 2007; Awiti et al., 2008). Islam
308 et al., (2003), reported similar results for CEC (r^2 = 0.64, RPD = 1.6), using a separate
309 validation set to test the performance of calibration models developed. However, these
310 accuracies can still be considered as good and useful for quantitative predictions.
311
312 **3.2 Comparison of laboratory and on-line measurement accuracy**
313 In order to compare the performance of the general calibration models for the
314 prediction of studied soil properties between laboratory and on-line scanned spectra,
315 the independent validation set (25 samples) collected from the field 1 in Duck End farm
316 was used. Generally, smaller accuracies are observed when using soil spectra
317 collected under on-line measurement conditions, as compared to that for spectra
318 collected under laboratory non-mobile scanning conditions (Table 7). Results confirm
319 that the predictions of the laboratory and on-line measurements were classified as
320 excellent/very good for pH (RPD = 2.69 and 2.14 and r^2 = 0.86 and 0.78, respectively),

321 and moderately good for CEC (RPD = 1.77 and 1.61 and r^2 = 0.68 and 0.62,
322 respectively) and Mg_{ex} (RPD = 1.72 and 1.49 and r^2 = 0.66 and 0.67, respectively). For
323 Ca_{ex}, very good accuracy was calculated for laboratory method (RPD = 2.19 and r^2 =
324 0.86), as compared to the poor accuracy for the on-line method (RPD = 1.30 and r^2 =
325 0.61).
326 There is little literature about the on-line prediction of secondary soil properties e.g.
327 properties without direct spectral responses in the NIR spectroscopy. Soil pH was the
328 most successfully measured property with on-line vis-NIR sensors, with all reports
329 show less accurate on-line predictions than the ones achieved in the current study (r^2 =
330 0.78) with r^2 values of 0.62 (Christy, 2008) and 0.61 (Shibusawa et al., 2001). Only
331 comparison between measured and predicted pH maps was provided by Mouazen et
332 al. (2007), showing similar trends of spatial distribution. The low prediction accuracy
333 obtained in this study for on-line Ca_{ex} might be attributed to the uneven distribution of
334 sample concentration over the entire concentration range. In spite of the wide range on
335 concentration, the majority of samples have low values for Ca_{ex} while there is a small
336 group of samples with very high values (Fig. 4). Thus, there is a gap of values between
337 both groups, hindering the creation of a robust calibration model. Similar comment can
338 be made for laboratory scanned spectra (data not shown). Similar trend is observed for
339 Mg_{ex} with a lower impact on the accuracy for both laboratory non-mobile and on-line
340 measurement. To reduce this effect samples from Ely field (UK), with the highest
341 values for Ca_{ex} and Mg_{ex}, were removed during the development of calibration models,
342 which led to the results shown in Table 7 and Figure 4. In the future, it is recommended
343 to consider new samples with Ca_{ex} and Mg_{ex} values covering the gap in the
344 concentration range for the development of more robust calibration models. The low
345 accuracy in predictions of the Ca_{ex} model could also be related with the extraction
346 method selected. This method might be inappropriate for Ca_{ex} detection in
347 predominantly calcareous samples.
348

3.3 Analysis of error

3.3.1 Histogram of error

The histogram of normal distribution plots of error was calculated by subtracting predicted from measured values for each property using 25 samples of the independent validation set (Tabl 8). It can be clearly observed that the mean, variance and SD values are much larger with laboratory scanned than with on-line scanned spectra. This is a clear sign to support the conclusion that accuracy decreases during on-line measurement due to the ambient conditions (e.g. noise, vibration, stones, plant roots, and a spectrum position with a corresponding soil sample) (Mouazen et al., 2007; Stenberg et al., 2010). However, under laboratory scanning conditions all these ambient conditions affecting accuracy are eleminated.

3.3.1.1 Histograms of error for pH

The histogram plot of normal error distribution for pH prediction using laboratory scanned spectra is normally distributed around 0 (Figure 5a), with slight skewness (0.392) towards the positive side, indicating slight overestimation. However, for on-line prediction, a smaller skewness (0.024) can be observed (Table 8). Around 76 % and 64 % of laboratory and on-line predictions, respectively, are with smaller error than 0.3 in absolute values, which is smaller than 12 % of the normal pH range for agricultural soils (4 - 9) (USDA, 1998). Furthermore, the range of prediction error is larger for on-line prediction (-1.003 to 1.122), as compared to laboratory prediction. The smaller range of prediction error in addition to the larger portion of small error than 0.3 (76 %) of laboratory scanned spectra suggests the better performance of the general pH model for prediction based on laboratory than on-line scanned spectra.

372 **3.3.1.2 Histograms of error for CEC**

373 Similar to pH, the histogram plots of CEC error for both laboratory and on-line
374 predictions show normal distributed around 0 (Figure 5b), with skewness values of -
375 0.004 and -1.201, respectively (Table 8). Around 64 % and 76 % of laboratory and on-
376 line predictions, respectively, are with smaller error than 1.1 cmol+/kg, which are
377 smaller than 4.4 % of the normal CEC range for agricultural soils (CEC = 0 cmol+/kg
378 for sandy soils to about 50 cmol+/kg for clay soils (Mengel and Kirkby, 1982). This
379 surprisingly shows a bigger proportion of error around 0 for on-line prediction, as
380 compared the laboratory prediction. However, the range of error was larger for on-line
381 prediction than for laboratory prediction (-3.908 to 2.483 cmol+/kg). This might be
382 attributed to difficulties associated with matching position of a soil sample collected for
383 laboratory chemical analysis and the corresponding spectrum collected during on-line
384 measurement (Mouazen et al., 2007).

385 **3.3.1.3 Histograms of error for Ca_{ex}**

386 The normal distribution plot of error for Ca_{ex} prediction using laboratory scanned
387 spectra (Figure 5c) is clearly skewed towards the negative side of the plot. This
388 skewness is larger for on-line (skewness = -0.523) than for laboratory (skewness = -
389 0.010) predictions (Figure 5c, Table 8). About 80 % and 52 % of laboratory and on-line
390 predictions, respectively, are with smaller error than 5.5 cmol+/kg in absolute values,
391 which are smaller than about 9 % of the Ca_{ex} range present in the samples. This
392 confirms that laboratory prediction of Ca_{ex} is more accurate than the on-line prediction,
393 which is supported by a larger range of error for the on-line, as compared to the
394 laboratory prediction (-21.911 to 9.025 cmol+/kg).

3.3.1.4 Histograms of error for Mg_{ex}

The normal distribution plot of error of Mg_{ex} of laboratory predictions presents a rather normal distribution of error around 0. However, this is clearly skewed towards the negative range when on-line spectra are used (Figure 5d, Table 8). About 56 % and 48 % of laboratory and on-line predictions, respectively, are with smaller error than 0.17 cmol+/kg in absolute values, which are a smaller error than about 4.7 % of the Mg_{ex} range present in the samples. This indicates that as for pH and Ca_{ex}, the prediction of Mg_{ex} is more accurate and stable when laboratory scanned spectra are used, as compared to on-line measurement. This did not prove clearly for CEC, although the trend is similar to the other three properties.

The above discussion about histogram plots of error and accuracy analysis confirms that calibration models developed with laboratory scanned spectra perform better when validated with laboratory measured spectra, as compared to on-line measured spectra. However, in comparison with laboratory chemical and vis-NIR analyses, the on-line measurement of soil properties enables the collection of high number of data points (around 12.800 readings for 21 ha of the field 1), with average of around 2 points per metre travel distance. This large amount of data allows the spatial interpolation to estimates values for un-sampled points in the field. Then, the possibility to predict several soil properties from the same spectrum, the opportunity to create maps from this large amount of information and its utilisation for site specific land management within the field, suggest the on-line measurement of soil properties as a valuable measuring technique.

4 Conclusions

This paper reports on the performance of a vis-NIR spectroscopy-based on-line sensor for the prediction of soil properties without direct spectral response in the NIR range, namely, pH, CEC, Ca_{ex} and Mg_{ex}. It also compares prediction accuracy of these

properties between on-line and non-mobile laboratory scanning. The results obtained in this study allow the following conclusions to be drawn:

1- The on-line measurement system enabled the simultaneous measurement of several soil properties without direct spectral responses in the NIR spectroscopy across a field.

2- General calibration models, developed with samples collected from fields in different European counties are a successful procedure for the calibration of the on-line vis-NIR sensor for the prediction of pH, CEC and Ca_{ex} and Mg_{ex}.

3- Higher accuracy was obtained for predictions using laboratory scanned, as compared to on-line scanned spectra. The laboratory and on-line predictions were classified as excellent/very good for pH (RPD = 2.69 and 2.14 and r^2 = 0.86 and 0.78, respectively), and moderately good for CEC (RPD = 1.77 and 1.61 and r^2 = 0.68 and 0.62, respectively) and Mg_{ex} (RPD = 1.72 and 1.49 and r^2 = 0.66 and 0.67, respectively). For Ca_{ex}, very good accuracy was calculated for laboratory method (RPD = 2.19 and r^2 = 0.86), as compared to the poor accuracy for the on-line method (RPD = 1.30 and r^2 = 0.61).

4- The histogram plots of error proved the general calibration models developed with laboratory scanned spectra to perform better when used to predict the studied soil properties using laboratory scanned spectra, as compared to on-line scanned spectra.

5- The ability of continuous data gathering with the on-line soil sensor at a relatively high operational speed (about 3 km/h) with very good to moderate accuracy obtained for some of the properties investigated (e.g. pH, CEC and Mg_{ex}), suggest the recommendation of the on-line soil sensor for site specific fertilisation.

Further research is needed to upgrade the calibration models of pH, CEC, Ca_{ex} and Mg_{ex}, developed in this study using samples collected from a larger number of fields and countries with even distribution of concentrations along the entire concentration

449 range encountered in agricultural soils. This is recommended to improve the prediction
450 accuracy and robustness of models developed for studied soil properties.
451
452 **References:**
453
454 Awiti, A. O., Walsh, M. G., Shepherd, K. D., and Kinyamario, J. (2008). Soil condition
455 classification using infrared spectroscopy: A proposition for assessment of soil
456 condition along a tropical forest-cropland chronosequence. Geoderma, 143, 73–84.

457 Ben-Dor, E., and Banin, A. (1995). Near-infrared analysis as a rapid method to
458 simultaneously evaluate several soil properties. Soil Sci. Soc. Am. J., 59, 364–
459 372Bricklemyer, R. S., Brown, D. J. (2010). On-the-go VisNIR: Potential and limitations
460 for mapping soil clay and organic carbon. Computers and Electronics in Agriculture, 70,
461 209-216.

462 British Standard, BS ISO 10390., (2005). Determination of Ph.

463 British Standard, BS 7755 Section 3.12., (1996). Determination of the potential cation
464 exchange capacity and exchangeable cations using barium chloride buffered at pH=
465 8.1.

466 Brown, D. J., Bricklemyer, R. S., and Miller, P. R. (2005). Validation requirements for
467 diffuse reflectance soil characterization models with a case study of vis–NIR soil C
468 prediction in Montana. Geoderma 129, 251–267.

469 Chang, C., Laird, D.A., Mausbach, M.J., Hurburgh Jr., C.R. (2001). Near-Infrared
470 Reflectance Spectroscopy-Principal Components Regression Analyses of soil
471 properties. Soil Science Society of America Journal 65, 480–490.

472 Chodak, M., Khanna, P., Horvarth, B., Beese, F. (2004). Near infrared spectroscopy for
473 determination of total and exchangeable cations in geologically heterogeneous forest
474 soils. Journal of Near Infrared Spectroscopy 12, 315–324.

475 Christy, C. D. (2008). Real-time measurement of soil attributes using on-the-go near
476 infrared reflectance spectroscopy. Comput. Electron. Agric., 61, 10–19.
477 Cohen, M. J., Prenger, J. P., and DeBusk, W. F. (2005). Visible-near infrared
478 reflectance spectroscopy for rapid, nondestructive assessment of wetland soil quality.
479 Journal of Environmental Quality, 34, 1422-1434.
480 Cozzolino, D., and Moron, A. (2003). The potential of near-infrared reflectance
481 spectroscopy to analyze soil chemical and physical characteristics. Journal of
482 Agricultural Science, 140, 65-71.
483 Dunn, B.W., Beecher, H.G., Batten, G.D., Ciavarella, S. (2002). The potential of near-
484 infrared reflectance for soil analysis – a case study from the Riverine Plain of south-
485 eastern Australia. Australian Journal of Experimental Agriculture 42, 607–614.
486 Groenigen, J. W., Mutters, C. S., Horwath, W. R., Kessel, C. (2003). NIR and DRIFT-
487 MIR spectrometry of soils for predicting soil and crop parameters in a flooded field.
488 Plant Soil., 250, 155–165.
489 Islam, K., Stingh, B., McBratney. A. B. (2003). Simultaneous estimation of several soil
490 properties by ultra-violet, visible, and near-infrared reflectance spectroscopy. Australian
491 Journal of Soil Research 41, 1101–1114.
492 Knadel, M., Thomsen, A., Greve, M. H. (2011). Multisensor on-the-go mapping of soil
493 organic carbon content. Soil Science Society of America Journal, 75:1799-1806
494 Kuang, B., Mouazen, A. M. (2011). Calibration of visible and near infrared
495 spectroscopy for soil analysis at the field scale on three European farms. European
496 Journal of Soil Science, 62(4): 629-636.
497 Kuang, B., Mahmood, H. S., Quraishi, Z., Hoogmoed, W. B., Mouazen, A. M., van
498 Henten, E. J. (2012). Sensing soil properties in the laboratory, in situ, and on-line: a
499 review. In Donald Sparks, editors: Advances in Agronomy, 114, AGRON, UK:
500 Academic Press, 155-224.

501 Malley DF., Yesmin L., Wray D., Edwards S. (1999). Application of near-infrared
502 spectroscopy in analysis of soil mineral nutrients. Communications in Soil Science and
503 Plants Analysis 30, 999-1012.
504 Martens, H., Naes, T. (1989). Multivariate Calibration, second ed. John Wiley and Sons
505 Ltd, Chichester, UK, 419 pp.
506 Mengel, k., Kirkby, E. A., (1982). Principles of plant nutrition, third ed. International
507 Potash Institute Bern, Switzerland, 655 pp.
508 Mouazen, A.M., De Baerdemaeker, J. & Ramon, H. (2005). Towards development of
509 on-line soil moisture content sensor using a fibre-type NIR spectrophotometer. Soil &
510 Tillage Research, 80, 171–183.
511 Mouazen, A.M. (2006). Soil Survey Device. International publication published under
512 the patent cooperation treaty (PCT). World Intellectual Property Organization,
513 International Bureau. International Publication Number: WO2006/015463;
514 PCT/BE2005/000129; IPC: G01N21/00; G01N21/00.
515 Mouazen, A. M., and Ramon, H. (2006). Development of on-line measurement system
516 of bulk density based on on-line measured draught, depth and soil moisture content.
517 Soil and Tillage Research, 86, 218-229.
518 Mouazen, A. M., De Baerdemaeker, J., and Ramon, H. (2006). Effect of wavelength
519 range on the measurement accuracy of some selected soil constituents using visual-
520 near infrared spectroscopy. Journal of Near Infrared Spectroscopy, 14(3), 189-199.
521 Mouazen, A.M., Maleki, M.R., De Baerdemaeker J., Ramon, H., (2007). On-line
522 measurement of some selected soil properties using a VIS–NIR sensor. Soil and
523 Tillage Research, 93, 13-27.
524 Mouazen, A. M., Maleki, M. R., Cockx, L., Van Meirvenne, M., Van Holm, L. H. J.,
525 Merckx, R., De Baerdemaeker, J., Ramon, H., (2009), Optimum three-point linkage set
526 up for improving the quality of soil spectra and the accuracy of soil phosphorus
527 measured using an on-line visible and near infrared sensor. Soil and Tillage Research,
528 103, 144-152.

529 Munoz, J. D., Kravchenko, A. (2011). Soil Carbon mapping using on-the-go near
530 infrared spectroscopy, topography and aerial photographs. *Geoderma*, 166, 102-110.
531 Nicola B. M., Beullens K., Bobelyn E., Peirs A., Saeys W., Theron K. I., Lammertyn J.
532 (2007). A review_Nondestructive measurement of fruit and vegetable quality by means
533 of NIR spectroscopy. Postharvest Biology and Technology 46 99–118
534 Pirie, A., Singh, B., Islam, K. (2005). Ultra-violet, visible and near-infrared diffuse
535 reflectance spectroscopic techniques to predict several soil properties. Australian
536 Journal of Soil Research 43, 713–721.
537 Shepherd, K.D., Walsh, M.G. (2002). Development of reflectance spectral libraries for
538 characterization of soil properties. Soil Science Society of America Journal 66, 988–
539 998.
540 Shibusawa, S., Made Anom, S. W., Sato, H. P., and Sasao, A. (2001). Soil mapping
541 using the real-time soil spectrometer. In ‘‘ECPA 2001’’ (G. Gerenier and S. Blackmore,
542 Eds.),Vol. 2, pp. 485–490. agro Montpellier, Montpellier, France.
543 Shonk, J. L., Gaultney, L. D., Schulze, D. G., and Scoyoc, G. E. V. (1991).
544 Spectroscopic sensing of soil organic matter content. Trans. ASAE., 34, 1978–1984
545 Stenberg, B., Viscarra Rossel, R., Mouazen, A.M., Wetterlind, J. (2010). Visible and
546 near infrared spectroscopy in soil science. Advances in Agronomy, 107: 163-215.
547 Udelhoven, T., Emmerling, C., and Jarmer, T. (2003). Quantitative analysis of soil
548 chemical properties with diffuse reflectance spectrometry and partial least-square
549 regression: A feasibility study. Plant Soil 251, 319–329.
550 USDA Natural Resources Conservation Service (1998). Soil Quality Information Sheet
551 Soil Quality Indicators: pH.
552 Viscarra Rossel, R. A., Walvoort, D. J. J., McBratney, A. B., Janik, L. J., Skjemstad, J.
553 O. (2001). Proximal sensing of soil pH and lime requirement by mid infrared diffuse
554 reflectance spectroscopy. In: Grenier, G., Blackmore, S. (Eds.), ECPA, Third European
555 Conference on Precision Agriculture, vol. 1. Agro Montpellier, pp. 497 – 508.

556 Viscarra Rossel, R. A., Walvoort, D. J. J., McBratney, A. B., Janik, L.J., Skjemstad,
557 J.O. (2006a). Visible, near infrared, mid infrared or combined diffuse reflectance
558 spectroscopy for simultaneous assessment of various soil properties. Geoderma 131
559 59 –75
560 Viscarra Rossel, R. V., McGlyn R. N., McBratney, A. B. (2006b). Determining the
561 composition of mineral-organic mixes using UV–vis–NIR diffuse reflectance
562 spectroscopy. Geoderma 146, 403-411.
563 Viscarra Rossel, R. A. and Behrens, T. (2010). Using data mining to model and
564 interpret soil diffuse reflectance spectra. Geoderma, 158, 46-54.
565 Volkan A. Bilgili, H.M. van Es, F. Akbas, A. Durak, W.D. Hively. (2010). Visible-near
566 infrared reflectance spectroscopy for assessment of soil properties in a semi-arid area
567 of Turkey. Journal of Arid Environments, 74, 229–238
568 Wetterlind, J. Stenberg, B. Söderström, M. (2010). Increased sample point density in
569 farm soil mapping by local calibration of visible and near infrared prediction models.
570 Geoderma, 156, 152-160.
571 Zornoza, R. C. Guerrero, J. Mataix-Solera, Scow K.M., Arcenegui V., Mataix-Beneyto
572 J. (2008). Near infrared spectroscopy for determination of various physical, chemical
573 and biochemical properties in Mediterranean soils. Soil Biology & Biochemistry, 40
574 1923–1930.